Cry78Aa, a novel Bacillus thuringiensis insecticidal protein with activity against Laodelphax striatellus and Nilaparvata lugens

Article (Accepted Version)

Title: Cry78Aa a novel *Bacillus thuringiensis* insecticidal protein with activity against *Laodelphax striatellus* and *Nilaparvata lugens*.

Running title: Cry78Aa insecticidal Bt toxin

Authors: Yinglong Wang[1, 2], Yonglei Liu[2, 3], Jie Zhang[2], Neil Crickmore[4], Fuping Song[2], Jiguo Gao[1#], Changlong Shu[2#],

Authors' affiliations:

[1]School of Life Science, Northeast Agricultural University, Harbin 150030, P. R. China.

[2]State Key Laboratory for Biology of Plant Diseases and Insect Pests, Institute of Plant Protection, Chinese Academy of Agricultural Sciences, Beijing 100193, P. R. China

[3]State Key Laboratory of Medical Vector Surveillance and Pathogen Detection, Beilun Entry and Exit Inspection and Quarantine Bureau, Ningbo 305012, P. R. China

[4]School of Life Sciences, University of Sussex, Brighton, BN1 9QG, UK

[#]**Corresponding author:** Changlong Shu, E-mail: clshu@ippcaas.cn

[#]**Corresponding author:** Jiguo Gao, E-mail: gaojiguo1961@hotmail.com

Abstract:

Transgenic plants expressing insecticidal proteins originating from *Bacillus thuringiensis* (Bt) have successfully been used to control lepidopteran and coleopteran pests with chewing mouthparts. However, only a handful of Bt proteins have been identified with any bioactivity against sap sucking pests (Hemiptera) including aphids, whiteflies, plant bugs and planthoppers. A novel Bt insecticidal protein with significant toxicity against a hemipteran insect pest is described here. The gene encoding the 359 amino acid, 40.7 kDa protein was cloned from strain C9F1. After expression and purification of the toxin, its median lethal concentration (LC_{50}) values against *Laodelphax striatellus* and *Nilaparvata lugens* were determined as 6.89 μg/mL and 15.78 μg/mL respectively. Analysis of the toxin sequence revealed the presence of both Toxin_10 and Ricin_B_Lectin domains.

Keywords: Planthopper, Hemiptera, Insecticidal protein

Introduction:

Rice is one of the world's most important food crops and most people living in Asia depend on it for part of their staple food. The rice planthoppers with a sucking mouthpart, not only feed on the phloem sap of rice plants but also serve as a vector leading to virus infection which can cause serious yield loss (Heong and Hardy, 2009). The brown planthopper (*Nilaparvata lugens*), small brown planthopper (*Laodelphax striatellus*) and white back planthopper (*Sogatella furcifera*) are three main hemipteran pests of rice and seriously threaten rice production. Currently, planthopper control methods rely mainly on the application of chemical insecticides. Not only can these induce resistance in the pest but are accompanied by the unintended killing of the non-target organisms.

As *Bacillus thuringenesis* (Bt) and plants expressing Bt insecticidal proteins have been successfully applied in insect control (Palma et al., 2014a), many efforts have been carried out to develop rice planthopper specific Bt insecticidal proteins. Shao et al. used protein engineering to modify a lepidopteran-specific Cry1Ab toxin with known gut binding peptides to create a hybrid protein with limited activity against the brown planthopper *N. lugens* (Shao et al., 2016). Using a membrane feeding protocol (Wang et al., 2014) we had previously identified a number of Bt strains demonstrating some level of activity against *L. striatellus.* One of these strains, (1012) encoded two toxins, Cry64Ba and Cry64Ca, that were confirmed to have high toxicity against rice planthoppers (Liu et al., 2018). Another one of the strains identified in that screen (C9F1) was phenotypically distinct from the above stain and is the subject of this investigation.

Material and methods

Strains, plasmid and growth conditions.

The C9F1 (CGMCC10782) strain was isolated form soil collected from the BaiWangShan Forest Park in Beijing and preserved at Institute of Plant Protection, Chinese Academy of Agricultural Sciences (IPPCAAS), Beijing. Scanning electron microscopy and SDS-PAGE analysis of the spore-crystal mixture of C9F1 were conducted following the methods described by Shu et al. (Shu et al., 2007). For Q-Exactive Mass Spectrometry analysis crystals solubilized in sodium carbonate buffer were subjected to SDS-PAGE, bands were excised, combined and subjected to in gel digestion with trypsin. The resulting fragments were analysed on a Q Exactive™ Hybrid Quadrupole-Orbitrap Mass Spectrometer (Thermo Fisher, USA) and the data using MASCOT 2.6. *E. coli* DH5a was used for routine transformations, while *E. coli* Rosetta (DE3) was used for the expression of the cloned genes. All genes were introduced into pET-21b plasmid where they were fused to an N-terminal His tag. All *E. coli* strains were cultured in Luria–Bertani (LB) medium at 37°C. Bt strains were incubated at 30°C in 1/2 LB liquid medium or agar plates. The concentrations of ampicillin and chloramphenicol used for bacterial selection were 100 μg/mL and 50 μg/mL respectively.

Preparation of genomic DNA, sequencing and computational analysis. Genomic DNA of C9F1 was prepared as described by Song et al. (Song et al., 2003). Genome sequencing was performed on an Illumina HiSeq 2500 platform, using a paired-end genomic library (insert size 500 bp) strategy with read lengths of 125 bp. Clear reads were reassembled by SOAPdenovo (Luo et al., 2012). Protein coding sequences were predicted by GeneMark (Besemer et al., 2001). Protein coding sequences were annotated using Blastp (Altschul et al., 1997) with a local Bt insecticidal toxin

71 database. The local database of Bt toxin proteins was founded by available quaternary rank Cry
72 toxins protein sequences listed on the website maintained by the Bt delta-endotoxin nomenclature
73 committee (http://www.lifesci.sussex.-ac.uk/home/Neil_Crickmore/Bt/) (Crickmore et al., 1998).
74 Conserved domains were annotated using the InterPro database (Finn et al., 2017). Homology
75 modeling was used to generate the three-dimensional protein structure of Cry78Aa protein in the
76 SWISS-MODEL workspace (Biasini et al., 2014). Signal peptides were predicted using SignalP 4.1
77 (Petersen et al., 2011).
78 **Gene cloning of putative toxin genes.** Primers used for amplification of putative toxin genes were
79 designed based on the nucleotide sequences from the draft genome of C9F1 (Table 1). Sequences
80 (CTGGTGGACAGCAAATGGGTCGG for upstream primers and
81 GGTGCTCGAGTGCGGCCGCAAG for downstream primers) which were homologous to the
82 pET21b plasmid were added to the 5'-termini of those primers for seamless assembly cloning.
83 Reverse complementary sequences of the above were used as primers to linearize the pET-21b
84 plasmid by PCR. PCR products was amplified using PrimerSTAR DNA polymerase (TaKaRa,
85 China) in a PTC-100 Peltier Thermal Cycler (MJ Research, USA). PCR reactions were run as
86 follows: incubation at 94 °C for 3min then 30 cycles at 94 °C, 30 s; 55 °C, 30 s ;72 °C, 5 min with
87 a final extension at 72 °C for 10 min. The gene fragments and the linearized vector were recombined
88 with a seamless assembly cloning kit (Clonesmarter, USA) following the manufacturer's
89 instructions. Then the ligated products were introduced into *E. coli* DH5a and verified by 3730XL
90 DNA sequencer (Applied Biosystems, USA).
91 **Protein expression and purification**. A single colony of *E. coli* Rosetta (DE3), containing the

recombinant plasmid, was selected and cultured in LB medium at 37℃ until the optical density reached 0.6-0.8, then IPTG to a final concentration of 0.5 mmol/L was added, the temperature turned down to 25℃ and the cells cultured for additional 8 h. Bacterial cells were collected by centrifugation at 8000×*g* for 5min. The pellet was resuspended in 20 mmol/L Tris-HCl pH=8.0 and cells sonicated on ice water at 60 W for 5 min with 3s on, 5s off cycle. The supernatant was collected and passed through a Ni^{2+} column, eluted by gradient concentrations of imidazole. Buffer exchange was conducted by dialysis in Tris-HCl to remove imidazole. Proteins were analyzed by SDS-PAGE and the concentrations of solubilized proteins were determined by ImageJ (National Institutes of Health) using BSA as a standard.

Bioassay. *L. striatellus* was used for screening the toxicity of the purified proteins encoded by C9F1 candidate toxin genes. Proteins were added in liquid artificial diet at a concentration of 100 μg/mL and packaged in a membrane feeding system. After 6 days, dead insects were counted (Wang et al., 2014). The mortality of *L. striatellus* to different C9F1 proteins was analyzed using one-way ANOVA tests followed by Tukey's HSD tests with SPSS 21.0.

Two hemipteran insects *L. striatellus* and *N. lugens*, two lepidopteran insects *Helicoverpa armigera* and *Plutella xylostella*, a coleopteran insect *Colaphellus bowringi*, and an important predator *Chrysoperla sinica* were chosen for testing the toxicity of Cry78Aa, the methods of bioassay are referred to in the following papers (Li et al., 2014; Tabashnik BE, 1993; Wu K, 1999; Yan et al., 2009) . Protein concentrations of 60 and 600 μg/g were initially used for those insects tested using solid diet (*P. xylostella, H. armigera* and *C. sinica*) and at 60 and 600 μg/mL for those with liquid diets (*L. striatellus, N. lugens* and *C. bowringi*). If insecticidal activity was detected,

dose-response assays were used to establish an LC_{50} value, which was calculated using SPSS 21.0 with Probit analysis. Each treatment was repeated three times.

Result

Initial characterization of the C9F1 strain.

The spore and crystal mixture of C9F1 were examined under a scanning electron microscope and revealed small spherical crystals (Fig. 1A). Total protein of sporulated C9F1 was analyzed by SDS-PAGE and revealed one major protein of around 140kDa as well as other minor ones (Fig. 1B).

Draft genome sequence and gene annotation of C9F1 putative pesticidal proteins.

Using the Illumina sequencing platform a total of 6,422,579 nucleotide base pairs were generated, and were assembled to 610 scaffolds with a genome size 6.21 Mb. The number of predicted protein coding sequences was 6861. After screening these putative proteins against a local Bt pesticidal protein database, 8 full-length protein coding sequences were identified (Table 2). Two of these were highly similar to known Cry8 proteins while the other six showed only weak similarity to other known toxins. To establish whether or not these putative toxins were produced by the native Bt strain the bands obtained by SDS-PAGE (Fig 1B) were cut out, combined, and subjected to peptide mass fingerprinting following trypsin digestion. Analysis of the results identified peptides corresponding to proteins encoded by Gene_1, Gene_3, Gene_7 and Gene_8. The 140kDa band observed in Fig. 1B is consistent with that expected from the Cry8 proteins encoded by Gene_7 and Gene_8. It is less clear which bands in Fig. 1B are likely to be those encoded by Gene_1 and Gene_3.

Protein expression and bioassay of the putative toxin proteins.

Five out of the eight putative genes sequences were successfully cloned into pET-21b. All five genes

could be expressed in *E. coli* Rosetta (DE3) cells after induction by IPTG. After nickel-affinity chromatography the purified proteins were analyzed by SDS-PAGE (Fig. 2). All the gene products ran with sizes consistent with their predicted molecular weights (Table 2). Although peptides corresponding to the proteins encoded by Gene_1 and Gene_3 were detected in the spore/crystal mix of C9F1, proteins corresponding in size to the recombinant toxins do not appear to be heavily expressed in the native strain (Fig. 1B).

A discriminatory dose bioassay was performed against *L. striatellus* using 100 μg/mL of the purified recombinant proteins. Figure 3 shows that only the protein encoded by Gene_3 gave an activity significantly different ($P<0.001$) to that of the buffer-only control. Further assays established an LC_{50} value for this protein against *L. striatellus* as 6.89 μg/mL (95% CL 5.48-8.38) . The protein was tested against five additional insect species. Of these only *N. lugens* proved to be sensitive to this toxin with an LC_{50} of 15.78 μg/mL (95% CL 13.04-18.25). Less than 50% mortality was observed when *P. xylostella*, *H. armigera, C. bowringi* and *C. sinica* were exposed to a high dose 600 μg g^{-1}/mL^{-1} of the Gene_3 encoded protein, although some mortality/weight gain inhibition was observed with the former two insects at this dose (Table 3).

Molecular characterization of the hemipteran-active gene.

Gene_3 is 1080 bp long and encodes a polypeptide of 359 amino acids with a deduced molecular mass of 40.7 kDa. No signal peptide was identified. Two conserved domains named Ricin B lectin (IPR000772) and Toxin_10 (IPR008872) are located at residue positions 26-153 and 192-358 respectively (Fig. 4B). The Ricin B lectin domain is a subset of the β-trefoil Ricin B-like lectins domain (IPR035992) and includes those domains containing characteristic QxW motifs (Hazes,

155 1996). In the case of our toxin the QxW motifs exist as the known variant QxF. The Toxin_10
156 domain is associated with a number of insecticidal toxins including the BinA mosquitocidal toxin
157 from *Lysinibacillus sphaericus* and the Bt toxins Cry35, Cry36, and Cry49. In all four of these the
158 Toxin_10 domain is preceded by a β-trefoil Ricin B-like domain, which in the case of Cry35 also
159 contains the QxW motifs. Due to the similarity to these existing toxins, and the demonstration of
160 pesticidal activity, the protein encoded by Gene_3 was named Cry78Aa1 by the *Bacillus*
161 *thuringiensis* toxin nomenclature committee. Using Cry35Ab (PDB 4JP0) as the template, a model
162 was built of Cry78Aa (Fig. 4A, GMQE=0.59).

163 **Discussion**

164 Only Cry64Ba, Cry64Ca (Liu et al., 2018) and modified Cry1Ab (Shao et al., 2016) had previously
165 been confirmed as having high toxicity against rice planthoppers. The discovery of another toxin in
166 this study will hopefully increase the potential of being able to control these economically important
167 pests. The SDS-PAGE profile of C9F1 indicates that the main protein(s) expressed by this strain are
168 around 140 kDa in size and based on the genome sequence are most likely Cry8 toxins. These toxins
169 are normally reported as being active against coleopteran species, although peptides from Cry78Aa
170 were detected in the spore/crystal mix using mass spectrometry which could account for the activity
171 noted in the initial screen against *L. striatellus*. Our recombinant Cry78Aa protein showed high
172 toxicity to *L. striatellus* and *N. lugens*, and there was also some evidence of an effect against both
173 *P. xylostella* and *H. armigera*. In contrast the Cry64Ba and Cry64Ca hemipteran-active toxins that
174 we previously described had no activity against *P. xylostella* or any of the other
175 lepidopteran/coleopteran insects tested (Liu et al., 2018).

Analysis of the sequence of Cry78Aa suggests that it has an architecture very similar to the so-called Bin-like toxins (de Maagd et al., 2003). These are β-pore forming toxins containing an N-terminal β-trefoil domain, proposed to be involved in receptor binding, and a C-terminal Toxin_10 domain believed to be the actual pore-forming domain. The structure of the homologous Cry35Ab toxin has been solved (Kelker et al., 2014) revealing that the β-trefoil domain is structurally distinct from the Toxin_10 one and fitting the 'head and tail' model of other β-pore forming toxins with pesticidal activity (Berry and Crickmore, 2017).

Due to their specific feeding behavior, proteins used to control hemipteran pests should be presented in the phloem sap. Experiments have indicated that Bt protein expressed in rice can be ingested by *N. lugens* (Bernal CC, 2002). Recently, the Cry51Aa2 protein has been optimized via various strategies resulting in more than a 200-fold increase in insecticidal activity against *Lygus hesperus* (73 µg/mL to 0.3 µg/mL), and which when expressed in cotton, caused a 30-fold decrease of *Lygus* spp. compared to the native control during field trials (Baum et al., 2012) (Gowda et al., 2016). Previously, we have reported that a mixture of Cry64Ba and Cry64Ca showed high toxicity (2.14-3.15 µg/mL) against two rice planthoppers (Liu et al., 2018). A Cry-related protein with sequence similarity to Cry41Aa was examined against *Myzus persicae* and its LC_{50} calculated as 32.7 µg/mL (Palma et al., 2014b). Given the technical obstacles of controlling sap-sucking pests with Bt, the need to identify proteins with good hemipteran activity remains. Cry78Aa is such a protein and furthermore is active without the need for either in vitro activation or a 2^{nd} component. The toxin shows no activity against *C. sinica* which is an important predator found in a variety of crop systems including paddy fields. As a result Cry78Aa has significant potential for the future

control of rice planthoppers.

Accession number. The accession number of genes identified from C9F1 are as follows Gene_1, KY780621; Gene_2, KY780622; Gene_3, KY780623; Gene_4, KY780624; Gene_5, KY780625; Gene_6, KY780626; Gene_7, KY780627; Gene_8, KY780628.

Acknowledgments

This study was supported by National Key R&D Program of China (Grant 2017YFD0200400) and the National Science and Technology Major Project of China (Grant 2014ZX0800912B).

Compliance with ethical standards

The manuscript does not contain experiments using mammals and does not contain studies on humans.

Conflict of interest

The authors declare no competing interests.

References

Altschul, S. F., et al., 1997. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research. 25**,** 3389-3402.

Baum, J. A., et al., 2012. Cotton plants expressing a hemipteran-active Bacillus thuringiensis crystal protein impact the development and survival of Lygus hesperus (Hemiptera: Miridae) nymphs. J Econ Entomol. 105**,** 616-24.

Bernal CC, A. R., Cohen MB, 2002. Effect of rice lines transformed with Bacillus thuringiensis toxin genes on the brown planthopper and its predator Cyrtorhinus lividipennis. Entomologia Experimentalis Et Applicata. 102**,** 21-28.

Berry, C., Crickmore, N., 2017. Structural classification of insecticidal proteins - Towards an in silico characterisation of novel toxins. J Invertebr Pathol. 142**,** 16-22.

Besemer, J., et al., 2001. GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. Nucleic Acids Res. 29**,** 2607-18.

Biasini, M., et al., 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res. 42**,** W252-8.

Crickmore, N., et al., 1998. Revision of the nomenclature for the Bacillus thuringiensis pesticidal crystal proteins. Microbiol Mol Biol Rev. 62**,** 807-13.

de Maagd, R. A., et al., 2003. Structure, diversity, and evolution of protein toxins from spore-forming entomopathogenic bacteria. Annu Rev Genet. 37**,** 409-33.

Finn, R. D., et al., 2017. InterPro in 2017-beyond protein family and domain annotations. Nucleic Acids Res. 45**,** D190-D199.

Gowda, A., et al., 2016. A transgenic approach for controlling Lygus in cotton. Nat Commun. 7**,** 12213.

Hazes, B., 1996. The (QxW)3 domain: a flexible lectin scaffold. Protein Sci. 5**,** 1490-501.

Heong, K. L., Hardy, B., 2009. Planthoppers : new threats to the sustainability of intensive rice production systems in Asia. International Rice Research Institute (IRRI), Los Banos, Philippines.

Kelker, M. S., et al., 2014. Structural and biophysical characterization of Bacillus thuringiensis insecticidal proteins Cry34Ab1 and Cry35Ab1. PLoS One. 9**,** e112555.

Li, Y. H., et al., 2014. Use of an artificial diet system to study the toxicity of gut-active insecticidal compounds on larvae of the green lacewing Chrysoperla sinica. Biological Control. 69**,** 45-51.

Liu, Y., et al., 2018. Cry64Ba and Cry64Ca, Two ETX/MTX2-Type Bacillus thuringiensis Insecticidal Proteins Active against Hemipteran Pests. Appl Environ Microbiol. 84.

Luo, R., et al., 2012. SOAPdenovo2: an empirically improved memory-efficient short-read de novo assembler. Gigascience. 1**,** 18.

Palma, L., et al., 2014a. Bacillus thuringiensis Toxins: An Overview of Their Biocidal

Activity. Toxins. 6**,** 3296-3325.
Palma, L., et al., 2014b. Molecular and insecticidal characterization of a novel Cry-related protein from Bacillus thuringiensis toxic against Myzus persicae. Toxins (Basel). 6**,** 3144-56.
Petersen, T. N., et al., 2011. SignalP 4.0: discriminating signal peptides from transmembrane regions. Nat Methods. 8**,** 785-6.
Shao, E., et al., 2016. Loop replacements with gut-binding peptides in Cry1Ab domain II enhanced toxicity against the brown planthopper, Nilaparvata lugens (Stal). Sci Rep. 6**,** 20106.
Shu, C., et al., 2007. Improving toxicity of Bacillus thuringiensis strain contains the cry8Ca gene specific to Anomala corpulenta larvae. Curr Microbiol. 55**,** 492-6.
Song, F., et al., 2003. Identification of *cry1I*-type genes from *Bacillus thuringiensis* strains and characterization of a novel *cry1I*-type gene. Appl Environ Microbiol. 69**,** 5207-11.
Tabashnik BE, F. N., Chilcutt CF, Cushing NL, Johnson MW, 1993. Increasing efficiency of bioassays: evaluating resistance to Bacillus thuringiensis in diamondback moth (Lepidoptera: Plutellidae). Journal of Economic Entomology. 86**,** 635-644.
Wang, Z. X., et al., 2014. Use of a dietary exposure system for screening of insecticidal compounds for their toxicity to the planthopper Laodelphax striatellus. Insect Science. 21**,** 667-675.
Wu K, G. Y., Lv N, 1999. Geographic variation in susceptibility of Helicoverpa armigera (Lepidoptera: Noctuidae) to Bacillus thuringiensis insecticidal protein in China. Journal of Economic Entomology. 92**,** 23-178.
Yan, G., et al., 2009. An engineered Bacillus thuringiensis strain with insecticidal activity against Scarabaeidae (Anomala corpulenta) and Chrysomelidae (Leptinotarsa decemlineata and Colaphellus bowringi). Biotechnol Lett. 31**,** 697-703.

Figure legends:

Fig. 1 Scanning electron microscope (A) and SDS-PAGE analysis (B) of a spore and crystal mixture of C9F1.

Fig. 2 SDS-PAGE analysis of purified proteins encoded by candidate insecticidal genes from C9F1 expressed in *E. coli* Rosetta (DE3). M, protein marker (PageRuler Prestained Protein Ladder, Thermo); lane 1, Gene_1; lane 2, Gene_3; lane 3, Gene_4; lane 4, Gene_8; lane 5, Gene_6. Proteins running in the expected position are marked with arrows.

Fig. 3 Toxicity of purified proteins (100 μg/mL) encoded by C9F1 candidate insecticidal genes against *L. striatellus*. NC: Negative control (Tris-HCl Buffer only).

Fig. 4 Sequence analysis of Cry78Aa. A: Simulated spatial structure of the Cry78Aa, pink: α-helices; green: β-sheets; red: putative transmembrane segments. The structure was visualized using PyMOL. B: Gene structure display of insecticidal proteins showing a similar domain architecture as Cry78Aa.